Department of Information Technology
National Institute of Technology Karnataka, Surathkal.
May 2018

Project Report

On

Automatic Medical Image Annotation and Content-based Image Retrieval using Relevance Feedback

Submitted by

Manali Ashish Shah (15IT218)
Ria Kulshrestha (15IT231)
Shivani Shrivastava (15IT243)
Sneha Suresh Patil (15IT245)

VI Sem B.Tech

Under the Guidance of

Dr. Sowmya Kamath S

Dept. of Information Technology,
NITK, Surathkal

Abstract

This project implements classification of medical images, content-based medical image retrieval,and relevance feedback method for image retrieval for enhancing image classification and hence retrieval performance. For image classification our system uses Convolutional Neural Network(CNN). For content-based image retrieval, our retrieval system uses the results from the softmax classifier used in our CNN architecture.To enhance the performance further we combine our image retrieval system with relevance feedback mechanism in CNN.

Contents

1   Introduction

Image annotation is the task of attaching meta-data in form of caption or keywords to any image. The need for automatic annotation of medical images is ever-growing. Medical images like CT scans, MRIs, X-ray scans play a crucial role in modern medicine. They aid in early detection, diagnosis and treatment.

Content based Image Retrieval is the content(features) based search for similar images in a large database. Keyword based image retrieval when done, requires manual annotation which has many practical limitations due to the ever-growing size of the database. Also, the accuracy of keyword based retrieval system will be dependent on the precision of the annotation done manually, hence prone to more error.The content based image retrieval techniques addresses both these problems.

Content based image retrieval can be achieved using a CNN based model. CNN or ConvNet is a sub-class of Deep neural networks. They are complex feed-forward neural networks. They do not require feature extraction as they learn the features on their own while training. Also they have high accuracy and hence used very commonly for these problems(classification and recognition). A CNN model in conjunction with massive amount of data automates the annotation process as well as boosts the classification performance.

ImageClef 2009 dataset acts a benchmark for medical annotation.The image sets(train and test) are based on the IRMA project.The images of the dataset are fully classified radio graphs. It consists of 20,000 training images and 10,000 testing images categorized into 193 classes. The class label of the images are complete IRMA codes.

2 Literature Review

2.1 Background

Outcomes of Literature Review

[8] employs the method of k-nearest neighbors to retrieve information.The retrieval considers the euclidean distance for fetching the information.It considers all the documents are relevant.

[5] proposes the method of using the decision trees for retrieval that considers relevance feedback.The edges linking the nodes of the decision trees are given weights that are trained as per the feedback from the user.The features that are found in relevant images are given more weights than the ones not found.

[6] uses the technique of center symmetric linear binary patterns to extract texture feature from the images. The method suits well for X-Ray images since it has contrasting bright foreground over dark background region.

2.2 Identified Gaps

The technique for k-nearest neighbors doesn't consider the relevance or non-relevance of the document.It assumes that all the documents are relevant.

The pruning of the decision tree will affect the ratio of relevance and non-relevance in the dataset and hence will affect performance of the model.The merging of the nodes will also affect the ratio and hence the results.

An image region is said to be flat if it has a nearly uniform intensity that is the variance of the intensity values within the region is very low. The LBP feature is not robust on flat image areas since it is based on intensity differences. It also ignores the actual intensity level at the location it is computed on.

2.3 Problem Statement

Content-based medical image retrieval and relevance feedback method for image retrieval for enhancing image retrieval performance.

2.4   Objectives

- Performing image classification based on content of the images.

- Building an image retrieval system.

- Provide mechanism for relevance feedback.

- Improving classification performance based on the relevance feedback.

3 Methodology and Framework

3.1 Network Architecture

Our multi-layered Convolutional Neural Network (CNN) has 5 convolutional layers which are followed by 4 fully connected layers. Each convolutional layer is followed by a ReLu activation layer which is in turn followed by a max pooling layer. The dimensions of each layer is as follows:

- Input convolutional layer: 224x224 input tensor with 32 filter of 5x5x3 and outputs a 112x112x32 tensor after ReLu and max pooling.

- Second convolutional layer: 112x112 input tensor with 64 filter of 5x5x32 and outputs a 56x156x64 tensor after ReLu and max pooling.

- Third convolutional layer: 56x56 input tensor with 128 filter of 5x5x64 and outputs a 28x28x128 tensor after ReLu and max pooling.

- Fourth convolutional layer: 28x28 input tensor with 256 filter of 5x5x128 and outputs a 14x14x256 tensor after ReLu and max pooling.

- Fifth convolutional layer: 14x14 input tensor with 256 filter of 5x5x256 and outputs a 7x7x256 tensor after ReLu and max pooling.

  Fully Connected layers:

- First fully connected layer: 2048 nodes

- Second fully connected layer: 256 nodes

- Third fully connected layer: 64 nodes

- Fourth fully connected layer: 193 nodes

The output of the fourth fully connected layer is fed into the Softmax classifier.

3.2 Training method

Factors to be considered while training of the model include:

- Choice of optimizer: Adam Optimizer

- Choice of loss function: Cross-Entropy Loss

- Batch size: 100

- Number of epochs: 5

- Initialization of weights/ choice of checkpoint: Random initialization

- Learning Rate: 1e-4

The back propagation algorithm is used to compute the gradient with respect to the parameters of the model in order to achieve gradient based optimization, we have used an Adam optimizer.

Adam, an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments.

The Softmax classifier is used to finally predict the class label of the input images. It gives the probability of each class as the output. The target class has the highest probability. All the probabilities lie between 0-1 and the summation of probabilities of all the classes is equal to 1.

Cross Entropy is used to calculate the loss while training. Cross-entropy loss, or log loss, measures the performance of a classification model whose output is a probability value between 0 and 1. Cross-entropy loss increases as the predicted probability diverges from the actual label.

4 Implementation

4.1 Work Done

The project is developed using TensorFlow, which is a deep learning software library developed by Google. It is used for high performance numerical computation. The main advantage of using TensorFlow is its flexible architecture which allows easy deployment of computation across various platforms.

Firstly we re-organized the data so that each folder contains images of the class identified by its by folder name.Then we create a file named label.txt which contains the list of all such folders.

After that we read the list of folders from label.txt and extracts features like height, width, color space, channels, format, filename, class label along with the image itself and creates a file of the format '.tfrecord', which is simple record-oriented binary format commonly used in Tensorflow application. It also splits the dataset into two sets i.e training and testing. The data is shuffled so that there is a random distribution amongst the two sets.

The CNN is trained for 5 epochs. It is then used to classify the testing images. The results of the classification are then stored in a CSV file. The CSV file contains the predicted label and the image name.

For the retrieval system we use the CSV file as reference. When a user enters a particular class number then a list of all the images with the same class in the CSV file is retrieved.

For feedback we ask the user to enter the image name and the relevant class number. After receiving the feedback, we retrain the model by freezing the convolutional layers and only updating the weights of the fully connected layers.

4.2 Results and Analysis

Number of epochs : 3
Number of classes : 58
Batch Size : 120
Accuracy of the model : 40%

The metrics for evaluation include :
k-precision

k-recall

Since there is a unique occurence of query in the test set, the numerator of k-recall and k-precision will always be 1 if it has correctly been idefntified by the model else it will be 0.Also, If the image does not exist in the dataset, the value will be infinity.

For k=5,10, 20 and 40
Query Entered : 740028.png
k-precision for k=5 : 0.2
k-recall for k=5 : 0.33
k-precision for k=10 : 0.1
k-recall for k=10 : 0.2
k-precision for k=20 : 0.05
k-recall for k=20 : 0.09
k-precision for k=40 : 0.025
k-recall for k=40 : 0.04

Due to the low accuracy of the model, the performance metrics k-precision and k-recall are not giving satisfactory values.Also, since there is an unique occurence of the image in the dataset the numberator for the information retrieval metrics is always going to be 1 or 0.The denominator will increase as we increase the value of k. The accuracy of the CNN model can be improved by training it for more number of epochs. Since there is an unbalanced distribution of data among the classes, we have chosen the top 58 classes and have eliminated the classes with lower number of images to train the dataset.

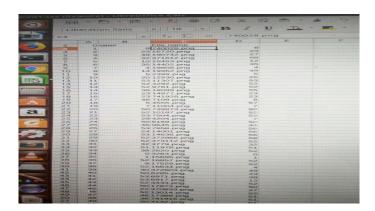Figure 1 shows the CSV file obtained for the testing images.

Figure 1: Result and Analysis

4.3   Innovative Work

The technique mentioned in the paper that considered the relevance feedback is via regular updation of the weights of the decision tree, that is giving more weights to the relevant features whereas ess weights to the non-relevant features. Our project uses Deep Learning CNN to decode the images and employs the relevance feedback by back propagating the error calculated during testing as per the feedback of the user. The back propagation updates the weight of the fully connected layers only, that is the initial layers of the model are freezed and not affected by the feedback. The results have been verified via printing the logit values before and after the feedback from user.

4.4   Details of each individual's work w.r.t. project tasks.

Manali Shah: Data preprocessing and training the model using CNN
Shivani Shrivastava: Relevance feedback mechanism by freezing all layers except fully connected layers during back propagation
Ria Kulshrestha: Relevance feedback mechanism by freezing all layers except fully connected layers during back propagation
Sneha Patil: Data preprocessing and training the model using CNN

5 Conclusion and Future Work

In this project, we carried out the retrieval and relevance feedback of the user by employing the Deepl Learning CNN architecture. The relevance feebdback was achieved by freezing the initial layers of the architecture and training only the fully connected layers after considering the feedback from the user.The obtained model was then tested for the batch of 120 test images. The accuracy of the model after training it for 3 epochs on the COCO dataset is 40%. The accuracy achieved is not satisfactory and can be imporved by further training the model for more number of epochs. The performance metrics were hence quite low due to the low accuracy of the CNN model and unique occurence of the query image in the test dataset.

We intend to do the following in the future:

- Train the model on a richer dataset.

- Train the proposed on different kinds of medical images.

- Eliminate the need of IRMA codes by using NLP.

References

[1] http://www.imageclef.org/2009/medanno

[2] https://www.sciencedirect.com/science/article/pii/S0925231208001355

[3] Ojala T, Pietikainen M, Maenpaa T: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24:971987, 2002

[4] Latex Template for NITK, Adithya Bhat, Dept. of Information Technology, NITK, https://github.com/adithyabhatkajake/latex-template-nitk

[5] MacArthur SD, Brodley CE, Shyu CR: Relevance feedback decision trees in content-based image retrieval. Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries: 6873, 2000.

[6] Ko BC, Kim SH, Nam JY: X-ray image classification using random forests with local wavelet-based CS-local binary patterns. JDigit Imaging, 2011

[7] Maria Tzelepi, Anastasios Tefas: Relevance Feedback in Deep Convolutional Neural Networks for Content Based Image Retrieval, 2016

[8] Marius Muja, David G. Lowe:Fast Approximate Nearest Neighbors,2009 WITH AUTOMATIC ALGORITHM CONFIGURATION